ELSEVIER

Agricultural and Forest Meteorology 104 (2000) 59–65

AGRICULTURAL AND FOREST METEOROLOGY

www.elsevier.com/locate/agrformet

Conditional sampling revisited

John M. Baker

USDA-ARS, Department of Soil, Water, and Climate, 1991 Upper Buford Circle, St. Paul, MN 55108, USA

Abstract

There is a continued need for simple, robust, yet accurate methods for measuring the surface/atmosphere exchange of a wide variety of trace gases and particulates. Conditional sampling is a relatively new method that has received increasing attention in recent years because it is related to theoretically attractive eddy covariance, but does not require a rapid response sensor for the covariate. It does require rapid measurement of the vertical wind speed, w, and sorting of sampled air into two separate lines based on the direction of w. As originally proposed, the flux was then calculated as $F=\beta\Delta C\sigma_w$, where ΔC is the mean difference in concentration between the upward and downward moving eddies, σ_w the standard deviation of the vertical wind speed, and β an empirical coefficient. Subsequent exposition showed that β was derivable from the statistics of joint Gaussian distribution, although field experiments have consistently found values in the range of 0.56 to 0.58, somewhat lower than the theoretical expectation of $\approx$0.62–0.63. Here, we reexamine the method, and show that if the flux is instead expressed as $F=b_1\sigma_w^2$, where b_1 is the regression-estimated slope of the concentration vs. wind speed relation, then it is exactly equivalent to eddy covariance. The aim of conditional sampling then becomes an estimation of b_1 as $\Delta C/\Delta W$. We show that this quantity has a consistent positive bias when samples are sorted simply into positive and negative excursions from mean w. Inclusion of a sampling deadband, symmetric about the mean w, improves the accuracy of the slope estimate and decreases its variance as well.

A potential problem with conditional sampling, regardless of which formulation may be used, is the effect of random measurement error (noise) in the wind speed measurement. We show that this introduces systematic errors into conditional sampling, while eddy covariance measurements are unaffected. Direct and indirect assessments indicate that these errors are too small to be significant for the sonic anemometer that we used, but it is probably wise for practitioners of the method to make certain that such is the case for the instruments used in their particular systems. We conclude that conditional sampling is a maturing method, with an increasing body of evidence indicating that the underlying relationships between scalar concentration and wind speed are sufficiently robust to support widespread use. Published by Elsevier Science B.V.

Keywords: Conditional sampling; Surface–atmosphere exchange; Flux measurement

1. Introduction

Turbulent flux measurement continues to pose challenges to those interested in surface–atmosphere exchange (Baldocchi et al., 1988). Eddy covariance is the most attractive method from a theoretical standpoint, but it requires coincident measurement of both vertical wind speed and the concentration of the scalar of interest at a frequency sufficient to encompass all eddies contributing to transport. There are only a few entities for which sensors with sufficient dynamic response exist, so there has long been interest in alternative approaches that might somehow allow the use of slower instruments to measure concentration.

Businger and Oncley (1989) examined sets of raw eddy covariance measurements of vertical wind speed,

* *E-mail address:* jbaker@soils.umn.edu (J.M. Baker)

0168-1923/00/$ – see front matter Published by Elsevier Science B.V.
PII: S0168-1923(00)00147-7

humidity, and temperature and noted that if the scalar observations of temperature and humidity were segregated according to the sign of the vertical wind speed, then the flux was proportional to the product of σ_w and the difference in concentration between the two sample bins:

$$F = \beta\sigma_w\Delta C \tag{1}$$

They found that the coefficient of proportionality, β, was relatively insensitive to stability and appeared to be equal to $\approx$0.6. Subsequently, Baker (1992) and Wyngaard and Moeng (1992) independently derived the underlying basis of the method, and Baker (1992) built and demonstrated a system for measuring fluxes of H_2O and CO_2 by conditional sampling, using a portable infrared gas analyzer. The method has since been tested and applied to a variety of constituent gases, including methane, nitrous oxide (Beverland et al., 1996), ozone (Katul et al., 1996), and pesticides (Majewski, 1993).

Briefly, the derivation presented by Baker (1992) began with the eddy covariance equation, in which the flux of any entrained scalar is given by the mean covariance of its concentration and the vertical wind speed, recognizing that correction must be made for density fluctuations due to concurrent transport of heat and water vapor (Webb et al., 1980).

$$F = \overline{w'C'} \tag{2}$$

From regression analysis, by definition the mean covariance of two variables is:

$$\overline{w'C'} = r_{wC}\sigma_w\sigma_c \tag{3}$$

However, this is only true if the 'independent' or 'predictor' variable (w in this case) is measured without error. We will proceed with the derivation and return to this potentially important point later. In the original derivation, Baker (1992) substituted for r_{wC} an estimator known as the biserial correlation coefficient, r_b (Pearson, 1910):

$$r_b = \frac{pq}{z}\frac{\Delta C}{\sigma_C} \tag{4}$$

where q and p are the relative proportions of observations in the major and minor classes if the data are segregated according to w, and z the area under the unit normal curve cut off at q. When this substitution is made, the estimate of the covariance becomes:

$$\overline{w'C'} = \frac{pq}{z}\Delta C\sigma_w \tag{5}$$

Comparison of eqs. (5) and (1) shows that β is thus equivalent to the statistical estimator pq/z. The theoretical value of pq/z, when p and q are approximately equal, is in the range of 0.62 of 0.627, but empirical estimates of β based on raw eddy covariance data sets have always yielded lower numbers, in the vicinity of 0.56–0.58 (Baker, 1992; Pattey et al., 1993; Beverland et al., 1996; Katul et al., 1996). Katul et al. (1996) found that estimates of β varied among scalars during any given sampling period, but found that the mean values for each constituent over all time periods were not significantly different, all falling in the previously mentioned range of 0.56–0.58.

We now approach the issue in a slightly different way. By definition,

$$\overline{w'C'} = b_1\sigma_w^2 \tag{6}$$

where b_1 is the regression-estimated slope of C against w. Conditional sampling provides a simple means to approximate b_1 as $\Delta\bar{C}/\bar{w}$. The numerator can be obtained as it has been in previous implementations, from the mean concentration difference between the air streams sampled from the upward and downward eddies. The denominator is calculated from the means of the instantaneous velocities of the upward and downward eddies, and σ_w^2 the variance of the vertical wind speed, so that the flux is estimated as:

$$\overline{w'C'} = \frac{\Delta\bar{C}}{\Delta\bar{w}}\sigma_w^2 \tag{7}$$

This is not a radical departure from the original concept of conditional sampling, but there is a key difference: we are using some additional information that we previously threw away, namely the difference in mean wind speed between the upward and downward eddies. Inspection of Eq. (7) shows that this reduces to the same approach that was taken by Pattey et al. (1993), who used Eq. (1) but calculated β for each time period as $\sigma_w/\Delta\bar{w}$.

The connection between eddy covariance and conditional sampling is more evident from eqs. (6) and (7) than in the original derivation, Eq. (1). There is no empiricism in the form of a β coefficient, and there

is no reliance on biserial correlation. However there is an obvious question about the accuracy of $\Delta C/\Delta W$ as an estimator of b_1. Furthermore, there is a yet unplucked nettle briefly alluded to earlier, the effect of measurement error in the predictor variable, w. How does it affect conditional sampling?

If w_1 is the measured vertical wind speed, each value is composed of the true instantaneous wind speed, w, and a measurement error component, ϕ:

$$w_1 = w + \phi \tag{8}$$

The variance that is measured becomes

$$\sigma_1^2 = \sigma_w^2 + \sigma_\phi^2 \tag{9}$$

Thus, the measured estimate of the variance of the vertical wind speed is a biased estimator, with a bias factor >1. What about estimates of the covariance and the regression parameters? An analysis by Velikanov (1965) shows that the effect of measurement errors in both the predictor and response variables (assuming that those errors are random and uncorrelated) is a systematic underestimate of r, the bias factor being

$$\frac{\sigma_w \sigma_c}{\sigma_{w1} \sigma_{c1}} \tag{10}$$

where the standard deviations in the numerator are the 'true', and those in the denominator the measured standard deviations, which incorporate both the true underlying variability and the measurement error. The covariance is unaffected by the *random* measurement error, an important attribute of flux measurement by eddy covariance. However, the regression estimate of the slope is affected. The bias factor is <1, and depends on the ratio of the standard deviation of the predictor 'error' component to the true underlying standard deviation of the predictor. The situation becomes more complicated if the measurement errors of either variable are correlated with the predictor, i.e. if the covariance of the vertical wind speed with either measurement error is nonzero (Draper and Smith, 1981), but this would be a problem for all eddy covariance-based methods. In general, however, if the measurement error is small relative to σ_w, the regression estimate of the slope will be relatively close. If such is not the case, the recommended statistical approach (Wald, 1940; Bartlett, 1949) is to separate the data into three equal parts based on the predictor variable (in our case, w), throw out the middle third, and compute the slope from the quotient of the difference of the means of the two remaining bins.

How then will measurement error in w affect conditional sampling? Recall that the original derivation involves substitution of a biserial estimator of r, Eq. (4), into a formula for the covariance, Eq. (3), resulting in Eq. (5), repeated here for convenience:

$$\overline{w'C'} = \frac{pq}{z} \Delta C \sigma_w \tag{11}$$

Since random errors in measurement of w will cause systematic overestimate of σ_w, this equation will overestimate the flux if the biserial estimate of r that it contains is unbiased.

In the revised approach to conditional sampling, the covariance, or flux, is shown to be equal to the product of the regression estimate of the slope and the variance of the vertical wind speed (Eq. (6)). Again, random measurement error in w will cause a systematic overestimate of σ_w^2. In the case of Eq. (6), that error is cancelled out since b_1 is the regression-estimated slope, which by definition contains σ_w^2 in the denominator, and hence has a compensatory bias factor. Thus, Eq. (7) will yield the correct flux only if the slope that is computed from conditional sampling ($\Delta C/\Delta W$) is an accurate estimator of b_1. The methods suggested for extracting the true slope from data sets in which both variables contain measurement errors (Wald, 1940; Bartlett, 1949) should produce overestimates of b_1 and, therefore, should overestimate the flux if used in Eq. (7), if the measurement error is sufficient to cause non-negligible error in σ_w^2.

We now raise two questions for consideration in the remainder of the manuscript:

1. How good is the conditional sampling approximation of b_1?
2. Is the contribution of random measurement error to σ_w^2 large enough to affect flux measurements by conditional sampling?

2. Materials and methods

A number of data sets were used to examine the questions raised in the preceding discussion. All were collected at the University of Minnesota's Rosemount Agricultural Experiment Station, located 24 km south of St. Paul (44°45′N, 93°05′W) using a 1-D sonic

anemometer, a 0.0127-mm diameter thermocouple, and an open-path krypton hygrometer (Campbell Scientific, Logan, UT). The sensible heat flux data were recorded above a bare field immediately after planting in June 1992, and were previously used by Baker (1992) for computing β, the coefficient of proportionality in Eq. (1). The latent heat flux data were collected during the summer of 1996 over an alfalfa canopy. The sampling rate in all cases was 10 Hz, and the height of measurement 2 m. Finally, the random measurement error of our sonic anemometer was measured by installing it inside a growth chamber (60 cm × 115 cm × 100 cm high). All chamber orifices were sealed, the chamber was turned off, and the anemometer was sampled at 10 Hz. The variance of the apparent vertical wind speed was then computed for several successive 30-min periods.

3. Results and discussion

If the contribution of measurement error, σ_ϕ^2, to the total variance in w in Eq. (9) is significant, then we would expect that it would have a proportionately larger influence when the total variance is small, and one manifestation of this should be a dependence of empirically-derived β (Eq. (1)) on the total variance. According to Baker (1992), β was calculated by regressing sensible heat fluxes against the product of σ_w and the difference in mean heat content, $\rho C_p T$, between upward and downward eddies. For each of the two large data sets they obtained a value of 0.56. The first of those data sets, measured over a bare soil with a broad range of sensible heat fluxes, was reanalyzed by plotting the ratio of the original dependent and independent variables against each corresponding measured variance in vertical wind speed (Fig. 1). This is, in effect, a plot of individually determined β values against total measured variance in w. Data from time periods in which the absolute value of the sensible heat flux was $<5\,\mathrm{W\,m^{-2}}$ were excluded. Though there is understandably much more scatter in the data from periods with low turbulence, there is no apparent dependence of β on σ_2, suggesting that the contribution of measurement error to the measured variance of w is sufficiently small so that much of the preceding discussion regarding its possible effects is moot, at least for this particular sonic anemometer. Additional confirmation of this was found in the null measurements made by placing the anemometer in a sealed, unventilated growth chamber. The measured variances for a series of 30-min averaging periods were all below $9.0\times10^{-6}\,\mathrm{m^2\,s^{-2}}$, which would be a trivial contribution to the total variance measured in the field, except under conditions of extreme stability. Furthermore, the similarity among β values found by

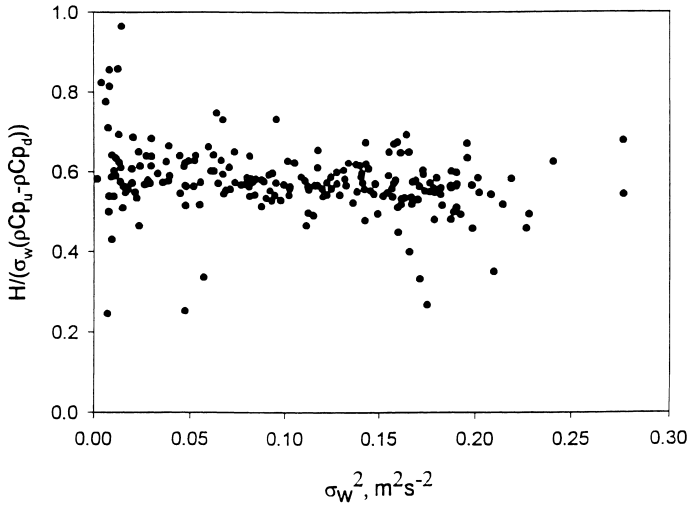

Fig. 1. Ratio of sensible heat flux to the product of σ_w and the difference in mean heat content between upward- and downward-moving eddies (β), plotted against variance in vertical wind speed.

independent groups, working with various different sonic anemometers, also supports the conclusion that, for commercially-available instruments, the contribution of measurement error to the total variance in w is negligible.

Why then are empirical determinations of β consistently somewhat lower than the theoretical expectation? The assumption that $\Delta\bar{C}/\bar{w}$ is a reasonable approximation for the regression estimate of the slope, b_1, is implicit within the original derivation of conditional sampling (Baker, 1992) and explicit within the revised derivation presented in this paper. Baker (1992) found that it overestimated the slope due to nonlinearity in the C vs. w relation, and hence was the primary source of deviation of β from its theoretical expectation. Katul et al. (1996) arrived at a similar conclusion based on eddy covariance data for four scalars. We examined this assumption further using a number of data sets containing 10-Hz measurements of vertical wind speed and absolute humidity, collected above an alfalfa canopy in the summer of 1996. The data were first analyzed to obtain the mean covariance and the regression parameters. Then they were processed as if they had been conditionally sampled; i.e. the data were sorted into bins according to the vertical wind speed. This was done for a variety of triggering strategies, simulating deadbands scaled to σ_w, and ranging from 20 to 100% of σ_w. After subdividing the data in this manner, mean values for wind speed and concentration in each bin were computed to determine

$$\frac{\Delta C}{\Delta W} = \frac{\bar{C}_{up} - \bar{C}_{dn}}{\bar{w}_{up} - \bar{w}_{dn}} \tag{11}$$

All slopes computed in this way were compared to b_1, the regression estimate of the slope, recalling that this value, when multiplied by σ_w^2, will by definition yield the covariance and, hence, the flux, regardless of whether the underlying relationship is truly linear.

Fig. 2 shows a typical set of results for one 30-min sampling period. Conditional sampling without a deadband overestimates b_1, Fig. 2A, which has a similar appearance to a plot for sensible heat shown by Katul et al. (1996). The estimate improves as the dead band increases, with the best estimate occurring at a deadband of approximately $\pm 0.9\sigma_w$ (Fig. 2B). Fig. 3 contains the same data from all sampling periods, and they show that conditional sampling without a deadband consistently produces $\Delta\bar{C}/\bar{w} > b_1$, with a mean ratio of b_1 to $\Delta C/\Delta W$ of 0.86. This is consistent with our earlier findings on both latent and sensible heat above a soybean canopy (Baker, 1992) and with the results of Katul et al. (1994). Furthermore, the effect of increasing deadband size on the accuracy of slope estimation is also consonant with the results of both Pattey et al. (1993) and Katul et al. (1996).

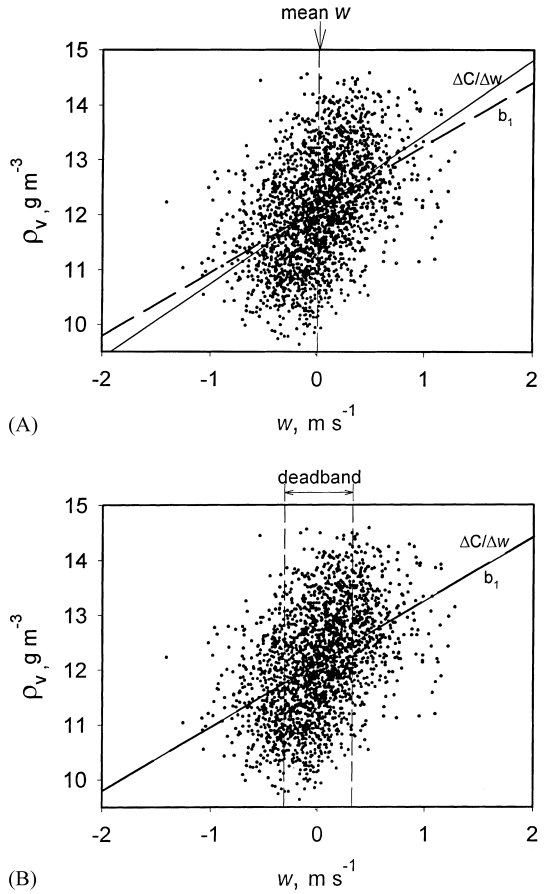

Fig. 2. Individual data points, sampled at 10 Hz above an alfalfa canopy. (A) Dashed line represents regression estimate of slope, b_1. Solid line represents slope estimated from $(\bar{C}_{up}-\bar{C}_{dn})/(\bar{w}_{up}-\bar{w}_{dn})$. (B) Same data as in (A). Again the dashed line is the regression estimated slope, b_1, but this time the solid line represents the slope estimated from conditional sampling with a deadband of $\pm 0.9\sigma_w$, i.e. $(\bar{C}_{up} - \bar{C}_{dn})/(\bar{w}_{up} - \bar{w}_{dn})$, calculated using only the data outside the vertical dashed lines.

Another consequence of using a deadband is that the separation between the means increases, i.e. $\Delta\bar{C}$ and $\Delta\bar{w}$ both become larger. This should have the effect

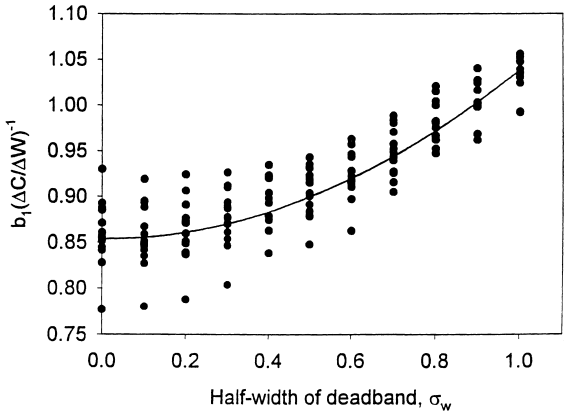

Fig. 3. Ratio of regression-estimated slope to conditional sampling estimate of slope, as a function of deadband width, for 10 separate 30-min intervals of 10 Hz sampling on vertical wind speed and vapor density. The data were actually sampled without a deadband; deadband widths >0 were simulated in post-processing. The fitted curve has the form $y=0.855+0.185x^2$.

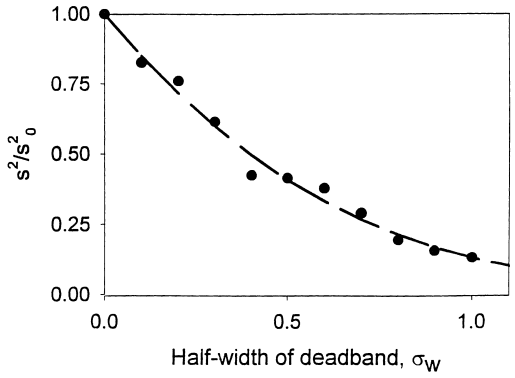

Fig. 4. Variance of the relative slope estimation, $b_1(\Delta\bar{C}/\Delta\bar{w})^{-1}$, as a function of deadband width. Dashed line is the theoretical expectation; the plotted points were calculated from the data shown in Fig. 3.

of reducing the variance in estimation of b_1, which is a function of the variance of the response variable and the sum of the squares of the residuals of the predictor variable:

$$\sigma_{b_1}^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{12}$$

This cannot be applied in any absolute sense for conditional sampling, but we can calculate a relative variance for the slope estimate as a function of deadband size for a normally distributed variable. That is shown as a dashed line in Fig. 4, along with the variance of the relative slope estimates at each deadband level in Fig. 3, also plotted against deadband size. The agreement is quite good, and leads to the following tentative conclusion: random errors in flux estimation by conditional sampling should decrease as deadband size is increased. There are practical limitations to this, since increasing the deadband size decreases the total amount of air sampled. Oncley et al. (1993) examined this issue and showed from theoretical considerations that the dependence of the 'signal-to-noise' ratio on deadband size should have a broad maximum in the vicinity of $\pm 0.6\sigma_w$.

The use of a deadband in conditional sampling thus seems to be a good idea. This should apply whether Eq. (1) or Eq. (7) is used, but with caveats. A deadband requires some additional plumbing and programming, both of which may introduce additional uncertainty into the determination of $\Delta\bar{C}/\bar{w}$. If a deadband is used, β in Eq. (1) will be different than it will be in the absence of a deadband, and the appropriate value can be calculated following Pattey et al. (1993). In Eq. (7), $\Delta\bar{C}/\bar{w}$ must be multiplied by a correction factor to obtain the best estimate of b_1, and that factor also depends on deadband width. An initial suggestion for the correction function is given in Fig. 3.

4. Conclusions

We have gently recast the derivation of conditional sampling to show that it is a method for estimating the covariance of vertical wind speed and scalar concentration (flux) as the product of the variance of the vertical wind speed and the regression slope, b_1, of the concentration vs. wind speed relation. There is a danger in this; random error in the wind speed measurement biases conditional sampling in ways that do not affect true eddy covariance. It appears that the noise level of commercially available sonic anemometers is sufficiently low that this is not significant, but it is probably necessary for potential practitioners to satisfy themselves that this is true for their particular instrument. We also found, consistent with previous investigations, that $\Delta\bar{C}/\bar{w}$ obtained from conditional sampling without a deadband overestimates b_1.

Inclusion of a deadband increases the accuracy and decreases the variance of slope estimation, with an optimum value in the vicinity of $\pm 0.9\sigma_w$.

The universality of conditional sampling remains an open question, but less so with each passing field study. The general consistency of results reported in the literature from diverse groups for a variety of transported scalars suggests that the joint distribution of wind speed and entrained species possesses robust features, encouraging the application of such relatively simple measurement methods to otherwise difficult environmental monitoring and research problems.

References

Baker, J.M., 1992. A field test of flux measurement by conditional sampling. Agric. Forest Meteorol. 62, 31–52.

Baldocchi, D.D., Hicks, B.B., Meyers, T.P., 1988. Measuring biosphere–atmosphere exchanges of biologically related gases with micrometeorological methods. Ecology 69, 1331–1340.

Bartlett, M.S., 1949. Fitting a straight line when both variables are subject to error. Biometrics 5, 207–212.

Beverland, I.J., O'Neill, D.H., Scott, S.L., Moncrieff, J.B., 1996. Design, construction, and operation of flux measurement systems using the conditional sampling technique. Atmos. Environ. 30, 3209–3220.

Businger, J.A., Oncley, S.P., 1989. Flux measurement with conditional sampling. J. Atmos. Ocean Tech. 7, 349–352.

Draper, N.R., Smith, H., 1981. Applied Regression Analysis. Wiley, New York. 708 pp.

Katul, G., Albertson, J., Chu, C.R., Parlange, M.B., Stricker, H., Tyler, S., 1994. Sensible and latent heat flux predictions using conditional sampling methods. Water Resour. Res. 30, 3053–3059.

Katul, G.G., Finkelstein, P.L., Clarke, J.F., Ellestad, T.G., 1996. An investigation of the conditional sampling method used to estimate fluxes of active, reactive, and passive scalars. J. Appl. Meteorol. 35, 1835–1845.

Majewski, M., 1993. Field comparison of an eddy accumulation and an aerodynamic-gradient system for measuring pesticide volatilization fluxes. Environ. Sci. Tech. 27, 121–128.

Oncley, S.P., Delany, A.C., Horst, T.W., Tans, P.P., 1993. Verification of flux measurement using relaxed eddy accumulation. Atmos. Environ. A 27, 2417–2426.

Pattey, E., Desjardins, R.L., Rochette, P., 1993. Accuracy of the relaxed eddy accumulation technique, evaluated using CO_2 flux measurements. Boundary Layer Meteorol. 66, 341–355.

Pearson, K., 1910. On a new method of determining correlation between a measured character A and a character B, of which only the percentage of cases where in B exceeds (or falls short of) a given intensity is recorded for each grade of A. Biometrika 7, 96–105.

Velikanov, M.A., 1965. Measurement Errors and Empirical Relations. US Department of Commerce, Springfield, VA, 230 pp.

Wald, A., 1940. The fitting of straight lines if both variables are subject to error. Ann. Math. Statist. 11, 284–300.

Webb, E.K., Pearman, G.I., Leuning, R., 1980. Correction of flux measurements for density effects due to heat and water-vapor density. Quart. J. R. Meteorol. Soc. 106, 85–100.

Wyngaard, J.C., Moeng, C.H., 1992. Parameterizing turbulent diffusion through the joint probability density. Boundary Layer Meteorol. 60, 1–13.